

Supporting Information

Rabeling *et al.* 10.1073/pnas.0806187105

SI Materials and Methods

Taxon and Gene Sampling. We extracted DNA from the right front leg, which we removed from the holotype, using a QIAGEN DNA Micro Kit following the instructions for small tissue samples. Two nuclear ribosomal genes and one single-copy nuclear gene were amplified using conventional PCR methods and sequenced. We obtained a total of 4926-bp-aligned (4197-bp-unaligned) sequence data: 1904 bp (1852 bp unaligned) for 18S; 2505 bp (1828 bp unaligned) for 28S; and 517 bp (517 bp unaligned) for elongation factor 1 F2. Discrepancies between aligned and unaligned numbers of base pairs are because of insertions in sequences of other taxa. We failed to sequence additional genes from the limited amount of extract. Primers are reported elsewhere (1, 2). The obtained DNA sequences were aligned with the data matrix provided by Brady *et al.* (1). DNA sequences are deposited at GenBank (www.ncbi.nlm.nih.gov/) as accessions EU913472–EU913474.

Phylogenetic Inference. The phylogenetic position of *Martialis* was estimated using both maximum likelihood (ML) and Bayesian approaches. Ambiguously aligned sites were excluded from both Bayesian and ML analyses. After excluding ambiguously aligned sites, the data set contained 1,048 variable sites, of which 765 were parsimony-informative characters.

Replicate, heuristic maximum likelihood (ML) searches in Garli v0.951 (3) were conducted until either the best-scoring topology was found twice or 50 replicate searches were performed. Generally, the two best topologies after 50 searches were topologically quite similar, if not identical. To ensure that ML scores were maximized, topologies resulting from all Garli searches were rescored in PAUP* (4). Analysis parameters were varied across Garli searches, although no set of parameters was found that consistently improved the search. Bootstrapping analyses used default analysis parameters with 500 bootstrap replicates.

Bayesian analyses were conducted with MrBayes v3.2 (5). (This is an unreleased version of MrBayes. The source code was downloaded from the current version system on Oct. 10, 2007). Characters were divided into four partitions: (i) the first and second codon positions of EF1 α F2, (ii) the third codon positions of EF1 α F2, (iii) the 18s rRNA gene, and (iv) the 28S rRNA gene. Parameters of the model of sequence evolution (i.e., relative rate matrix, base frequencies, transition/transversion ratio, proportion of invariant sites, and rate variation shape parameter) were unlinked across partitions, although topology, branch lengths, and rate multipliers remained linked. Model parameterization was chosen to reflect the models used by Brady *et al.* (1) and facilitate comparison to their results. Metropolis coupling was used with eight chains per analysis and a temperature increment of 0.05. Four independent replicates were run for each analysis. The state of the cold chain was recorded every 200 generations. Convergence was assessed using methods described in Brown and Lemmon (6) and implemented in the Java program, MrConverge v1b2 (A.R. Lemmon; available from <http://www.evotutor.org/MrConverge>). Briefly, these methods estimate the point of stationarity in likelihood scores, and the point at which the overall discordance in bipartition posterior probabilities is minimized. The latter of these two points in the chain is set as the burn-in. Convergence is reached when the widest 95% confidence interval for posterior probability across all bipartitions falls below a width of 0.1. At this point, the true bipartition posterior probability for all bipartitions lies within ± 0.05 of the estimated posterior probability.

Phyutility v2.2 (7) and FigTree v1.1.2 (8) were used to reroot trees, calculate consensus topologies, and assist in the creation of phylogenetic figures.

Topological Hypothesis Testing. Hypothesized rootings were compared using posterior probabilities, Bayes factors (BF; 9–12) and the Shimodaira–Hasegawa test (SH) (13, 14). ML and Bayesian analyses were used to estimate phylogenies under constraints corresponding to different rootings (Fig. 4), with the same analysis conditions as given above.

Marginal likelihoods from Bayesian analyses were estimated as the harmonic mean of the sampled likelihoods (10, 11, 15). The harmonic mean estimator has been found to perform well for phylogenetic applications (11). Bayes factors in favor of the *alternative* rooting were calculated as the ratio of marginal likelihoods, with the marginal likelihood of the alternative rooting in the numerator. The test statistic, $2\ln(\text{BF})$, was calculated with values less than -10 corresponding to strong support *against* the alternative (in favor of the maximum posterior probability rooting).

The SH test (13) was performed in PAUP* 4.0b10 (4) with RELL approximations and 1,000 bootstrap replicates, using the ML topology from the constrained searches under each of seven hypothesized rootings (Fig. 4). The SH test is a highly conservative test of topology (16).

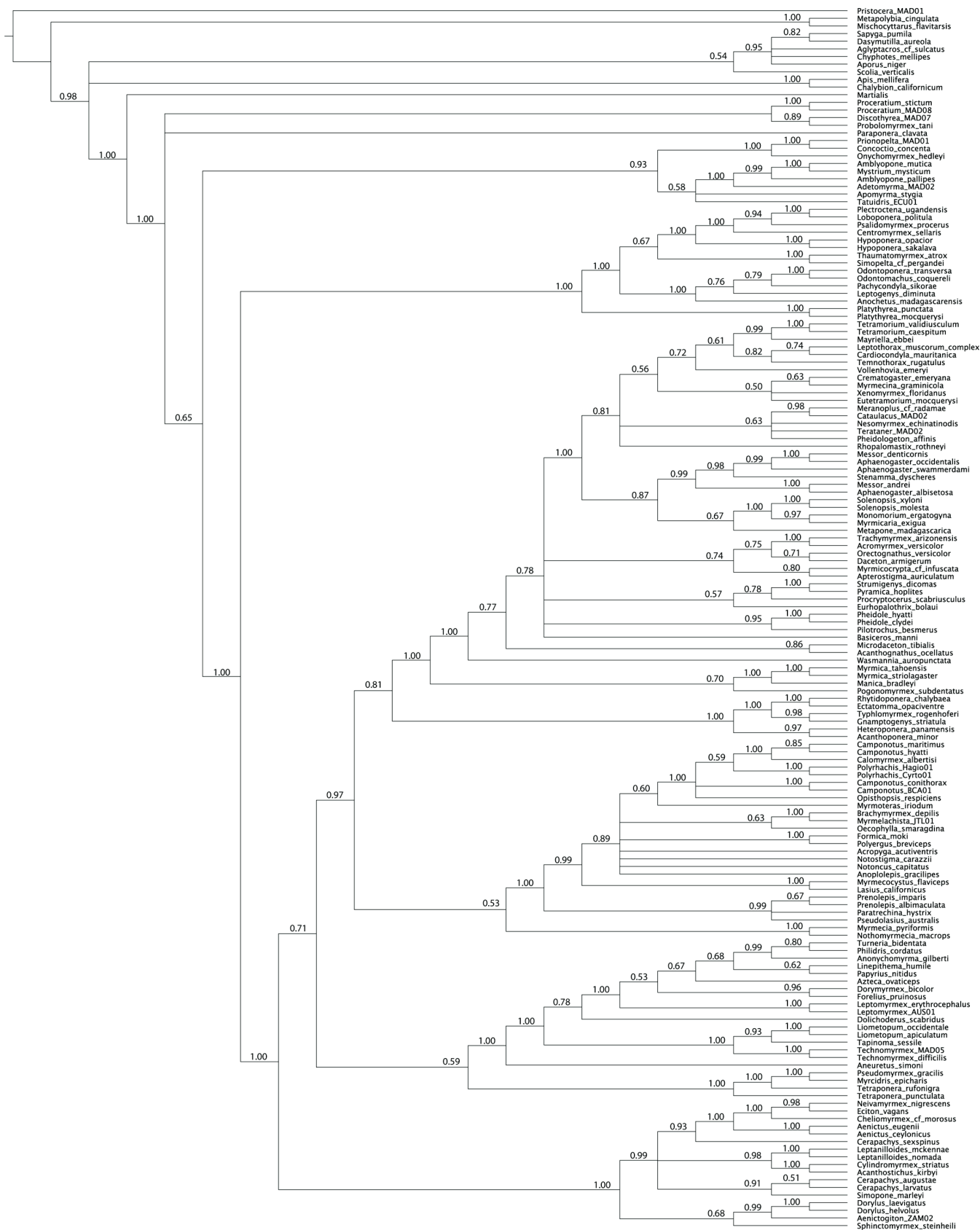
Model Adequacy Test. The adequacy of the model of sequence evolution used in the Bayesian analyses was assessed using posterior predictive simulation (17) with a multinomial likelihood test statistic, as implemented in the program PuMA v0.903 (18). One percent of posterior samples were used to simulate predictive data sets. The test statistic values of these simulated data sets defined the null distribution to which the test statistic value of the empirical data set was compared ([supporting information \(SI\) Fig. S3](#)).

We note that all phylogenetic estimates and topological hypothesis tests are contingent on the model of sequence evolution used in the analysis (16) although model adequacy is rarely evaluated in phylogenetic studies. Bayesian model adequacy tests reject the adequacy of the partitioned model used in our analyses ($P < 0.001$; [Fig. S3](#)). However, the magnitude and direction of any error induced by an inadequate model is unknown, thus we maintain that our results are still best possible estimates at this time. We do anticipate that future developments in phylogenetic methodology will produce adequate models of sequence evolution and our results and hypothesis should then be revisited.

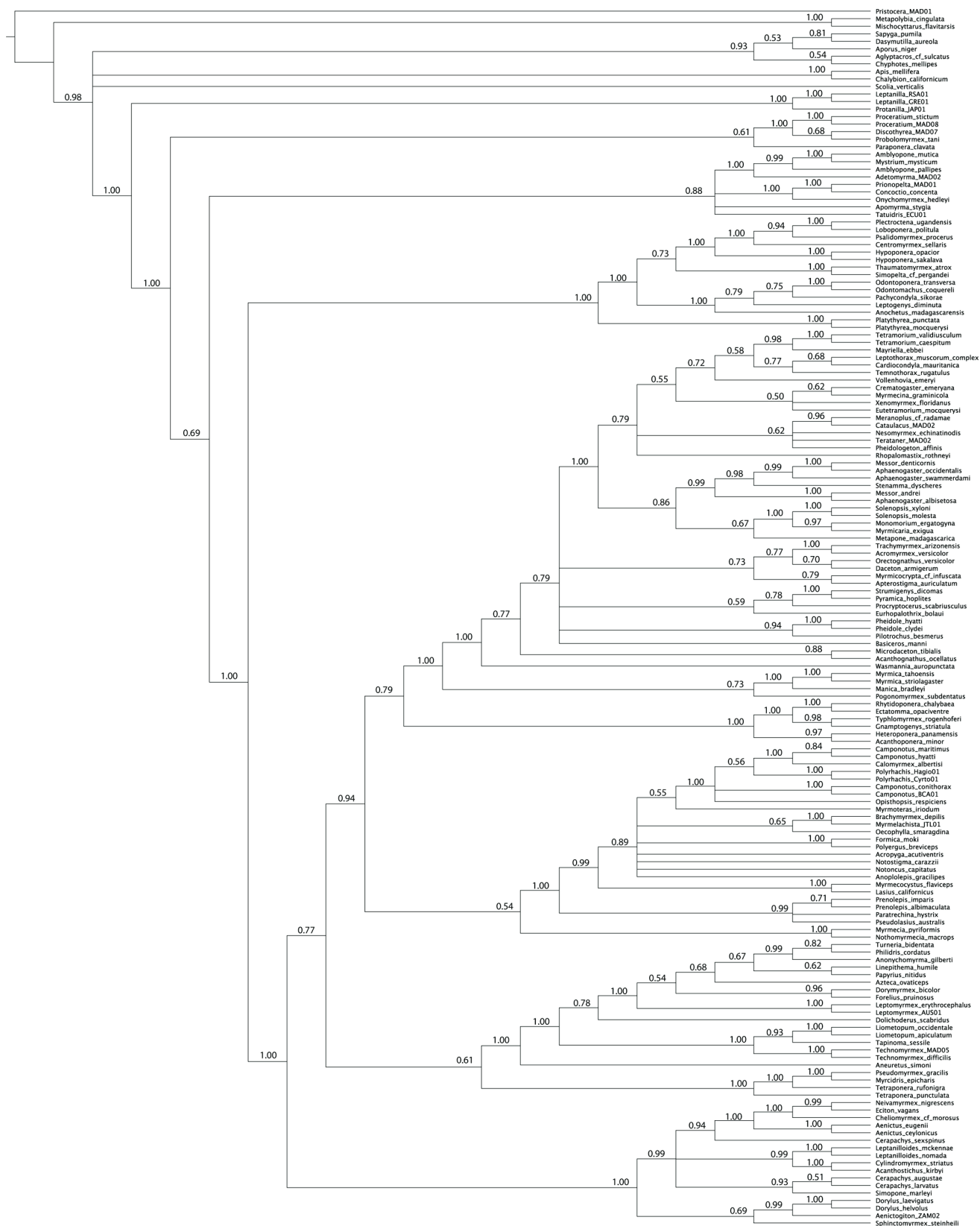
Morphometric Measurements. The only specimen available was examined using a Leica MZ16 stereo-microscope at magnifications of up to $184\times$. Measurements were completed at $115\times$ using a calibrated ocular micrometer. All metric measurements were recorded to the nearest 0.001 mm and rounded to two decimal places for presentation. Measurements are given in millimeters. Morphological terminology follows Bolton (19) and Dlussky (20). Anatomical abbreviations and definitions of measurements are shown in [Table S1](#).

Microscope Photography. Photomicrographs were taken using a JVC KY-F70 digital camera mounted on a Leica Z6 APO dissecting scope. Composite images were assembled from 20 sequential photographs with Syncroscopy AutoMontage (v 5.0) software (21).

1. Brady SG, Fisher BL, Schultz TR, Ward PS (2006) Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proc Natl Acad Sci USA* 103:18172–18177.
2. Ward PS, Downie DA (2005) The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): Phylogeny and evolution of big-eyed arboreal ants. *Syst Entomol* 30:310–335.
3. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation (Univer of Texas, Austin).
4. Swofford DL (2002) *PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods) v4.0b10* (Sinauer Associates, Sunderland, MA).
5. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
6. Brown JM, Lemmon AR (2007) The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol* 56:643–655.
7. Smith SA, Dunn CW (2008) Phyutility: A phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* 24: 715–716. Available at <http://code.google.com/p/phyutility>.
8. Rambaut A (2007) FigTree, a graphical viewer of phylogenetic trees. Available at <http://tree.bio.ed.ac.uk/software/figtree/>.
9. Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90:773–795.
10. Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Syst Biol* 53:47–67.
11. Brown JM, Lemmon AR (2007) The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol* 56:643–655.
12. Lavine M, Schervish MJ (1999) Bayes factors: What they are and what they are not. *Am Statist* 53:119–122.
13. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116.
14. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Syst Biol* 49:652–670.
15. Brandley MC, Schmitz A, Reeder TW (2005) Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol* 54:373–390.
16. Buckley TR (2002) Model misspecification and probabilistic tests of topology: Evidence from empirical data sets. *Syst Biol* 53:509–523.
17. Bollback JP (2002) Bayesian model adequacy and choice in phylogenetics. *Mol Biol Evol* 19:1171–1180.
18. Brown JM, Eldabaje R (2008) PuMA: Bayesian analysis of partitioned (and unpartitioned) model adequacy. Available at <http://code.google.com/p/phylo-puma>. Accessed March 1, 2008.
19. Bolton B (1994) *Identification Guide to the Ant Genera of the World* (Harvard Univ Press, Cambridge, MA).
20. Dlussky GM (1983) A new family of upper Cretaceous Hymenoptera: An intermediate link between the ants and scolioids. *Paleontol Zh* (1983)3:65–78 (translated from Russian; English translation in *Paleontol J* 17: 63–76).
21. Riedel A (2005) Digital imaging of beetles (Coleoptera) and other three-dimensional insects. *Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a Study of the European Network for Biodiversity Information*, Häuser et al. (Staatl. Museum für Naturkunde, Stuttgart). pp 222–250.







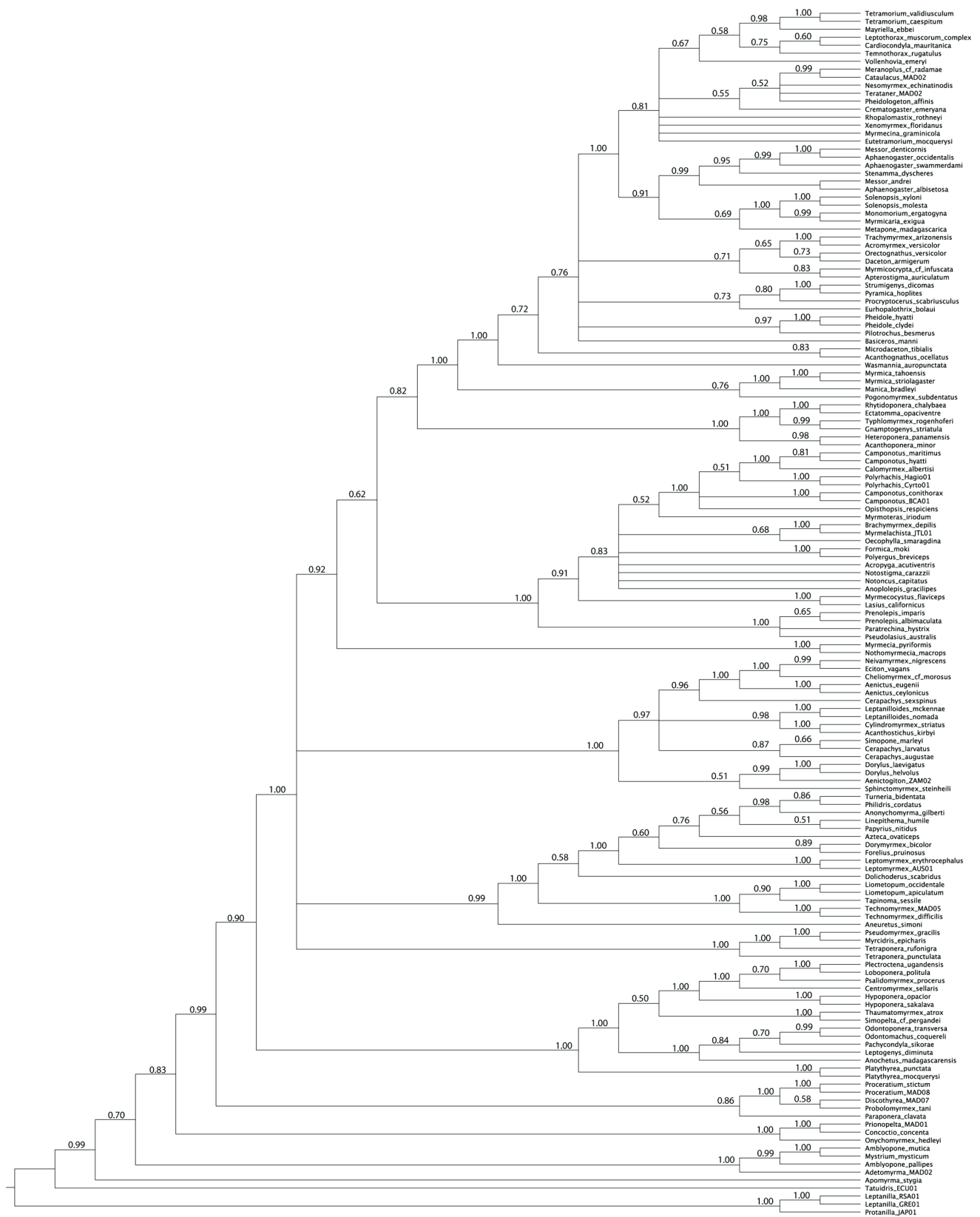


Fig. S1E.









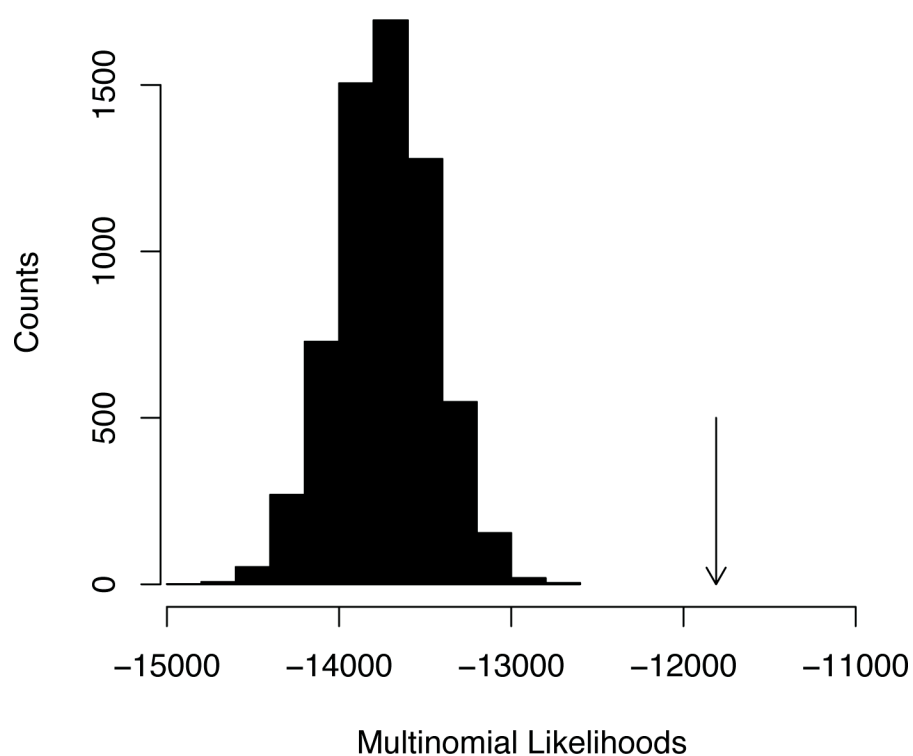


Fig. 53. Adequacy of the model of sequence evolution assessed with posterior predictive simulation. The null distribution was generated by subsampling (1%) output from the MCMC chains and using sampled parameter values to simulate data sets. The multinomial likelihoods of simulated data sets form the null distribution (histogram in black). The multinomial likelihood of the empirical data set is given by the arrow and clearly falls outside the null distribution ($P < 0.001$).

Table S1. Anatomical abbreviations and definitions of measurements

HL	Head length	Length of the head in full face view, excluding mandibles, measured in a straight line from the midpoint of the anterior clypeal margin to the midpoint of the posterior margin of the head, not including the projecting antennal sockets
HW	Head width	Maximum width of head
SL	Scape length	Maximum straight line length of the antennal scape excluding the basal constriction or neck close to the condylar bulb
FL	Funiculus length	Maximum length of antennal segments 2–12, excluding the scape
ML	Mandible length	Maximum mandible length, measured in a straight line from the mandibular insertion into the head capsule to the distal end of the mandible
WL	Weber's length	Diagonal length of the mesosoma in profile from the foremost point of the pronotum excluding cervical shield to the posterior base of the metapleuron
PW	Pronotum width	Maximum width of pronotum, measured in dorsal view
PEW	Petiolus width	Maximum width of petiolus, measured in dorsal view
PEL	Petiolus length	Maximum length of petiolus, measured in lateral view
PPW	Postpetiolus width	Maximum width of postpetiolus, measured in dorsal view
PPL	Postpetiolus length	Maximum length of postpetiolus, measured in lateral view
HFL	Hind femur length	Maximum length of metafemur
HTL	Hind tibia length	Maximum length of metatibia without the proximal part of the articulation which is received into the distal end of the hind femur
CI	Cephalic index	$HW*100/HL$
SI	Scape index	$SL*100/HW$
MI	Mandible index	$ML*100/HL$
DI	Dlussky index	$SL*100/FL$